

REMARKS

The Examiner has objected to the Abstract and a new abstract is enclosed herewith.

Claim 16 has been amended to depend from Claim 15.

The Examiner has rejected Claims 1-35 as being anticipated by Bregler. This rejection is respectfully traversed. Bregler system is based upon modification of the frames of an existing video. Applicants do not modify frames but analyze, transform, and generate, in continuous time, from any source to output to any media. Applicants' continuous time processing (see Page 14, lines 20-21 of the application) allows for 1 millisecond (or faster) accuracy in lip and facial movement, while Bregler's frame based system can do no better than 33 millisecond accuracy. Bregler states that a "phone" may encompass only three or four successive frames in the video, Col. 8, lines 28-30. Bregler's system relies entirely on phonemes to synchronize to a new script. Applicants use a symbol stream which is analyzed and processed at a 1 millisecond or faster rate versus Bregler's 33 millisecond rate. Bregler's frame rate precludes properly modeling some foreign languages and many facial expressions.

Bregler's symbol stream is comprised of phonemes et al at 10 millisecond acoustic frame rate (the de facto standard) and 2 dimensional visemes at a 33 millisecond image frame rate (the de facto standard). Applicants use a symbol stream comprised of continuous time coded facial and acoustic symbols at less than 1 millisecond acoustic frame rate and 2 dimensional visemes at less than a 1 millisecond image frame rate.

Bregler relies on speech recognition techniques to derive a phoneme stream. Applicants use many techniques to derive a high speed symbol stream including radar, lip reading, and a manual linguistic process, as well as speech recognition techniques.

In addition, Bregler tracks only the lips, Col. 7, lines 55-60. Applicants track and measure a number of control points in addition to the lips, including the chin, teeth, cheeks and jowls. This provides a much more detailed picture of the speaker's face for transformation. Bregler has no teaching of tracking these multiple locations nor any concept of how to do so.

Regarding Claim 2, Bregler mentions a dub in a different language but describes only the '532 patent as a method of doing so, Col. 1, lines 36-63. The '532 patent, as described uses only warping or morphing of the lips by manual output.

Regarding Claims 3 and 4, Applicants now point out that they employ a continuous time coded facial and acoustic symbol stream to animate the original speaker's face, not contemplated by Bregler.

Regarding Claim 5, the portion of Bregler cited by the Examiner refers to the lip tracking only of the '532 patent.

Regarding Claims 6 and 19, Bregler mentions a computer system but does not explain its use. Furthermore, Bregler is still limited to the lip tracking of phonemes.

Regarding Claim 7, the only fixed reference points contemplated by Bregler are the lips, done by a frame by frame analysis.

Regarding Claim 8, Bregler's database is of triphones and diphones measured by the frame analysis.

Regarding Claim 9, Col. 11, line 4 of Bregler does not mention an emotional elicitation by tracking a number of "facial control points". Bregler only describes tracking the lips speaking in phones, not a continuous tracking of 1 or less millisecond.

Regarding Claims 10, 20, 21 and 34, Bregler at Col. 1, line 49, only describes the tracking of the "outlines of the actor's lips".

Regarding Claims 11 and 22, the cited portion of Bregler only mentions “all lip movements”.

Regarding Claim 12, again the cited portion of Bregler only mentions recording all lip movements, not muzzle patches.

Regarding Claims 13, 14, 23 and 24, while Bregler mentions visual image sequences, they are solely related to the position of the lips to form various phonemes, Col. 5, lines 57-60.

Claim 18 has been cancelled.

In summary, Bregler uses phones as his highest resolution representation of speech (shortest time frame of 10 millisecond acoustic and 33 milliseconds video) as the Examiner noted in Col 4, line 40-41. Triphones and diphones are simply the set of phones (still at the same time frame) that describe the difference pairs from 1 phone to the next phone in speech. Applicants employ a much faster rate than Bregler and do so in a way that a much faster and thus more detailed rate can be accomplished, not simply the standard and fixed rates Bregler uses. In addition, triphones are not “the smallest units of speech” as Bregler states. They are simply the most common. Bregler’s visemes are recorded at 33 milliseconds video frames. Applicants record visemes in continuous time of 1 millisecond or less. Applicants’ smallest units of speech are much shorter in time, in order encompass rapid facial movement for example, not possible in Bregler’s system.

Applicants mark automatically the whole set of facial features, not just the lips. These include the chin, inner lips, outer lips, cheeks, jowls, teeth, nose and eyes. This is not contemplated by Bregler.

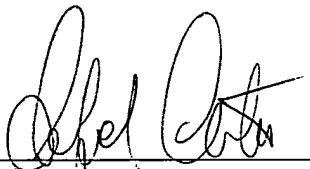
In one embodiment of Applicants’ invention, all of the measurements are tracked by radar, a feature not contemplated by any reference cited by the Examiner.

In another embodiment Applicants utilize a 3 dimensional muzzle model. Bregler works solely in 2 dimensions.

In view of the amendment of the Claims and the arguments above presented, early allowance of the claims is respectfully requested.

Respectfully submitted,

Birch, Stewart, Kolasch & Birch, LLP

by 
Sanford Astor
Attorneys for Applicants
310-209-4400